

MODULAR: Software for the Autonomous Computation of Modularity in Large Network Sets

Flavia Maria Darcie Marquitti^{1,*}, Paulo Roberto Guimaraes Jr.¹, Mathias Mistretta Pires¹, Luiz Fernando Bittencourt²

1 Departamento de Ecologia, Universidade de São Paulo, São Paulo, SP 05508-900, Brazil

2 Instituto de Computação, Universidade Estadual de Campinas, Av. Albert Einstein, 1251, Campinas, SP 13083-852, Brazil

* E-mail: flamarquitti@gmail.com

Abstract

Ecological systems can be seen as networks of interactions between individual, species, or habitat patches. A key feature of many ecological networks is their organization into modules, which are subsets of elements that are more connected to each other than to the other elements in the network. We introduce MODULAR to perform rapid and autonomous calculation of modularity in sets of networks. MODULAR reads a set of files with matrices or edge lists that represent unipartite or bipartite networks, and identify modules using two different modularity metrics that have been previously used in studies of ecological networks. To find the network partition that maximizes modularity, the software offers five optimization methods to the user. We also included two of the most common null models that are used in studies of ecological networks to verify how the modularity found by the maximization of each metric differs from a theoretical benchmark.

Introduction

Ecological systems can be seen as networks in which the elements, such as habitat patches within a landscape or species within communities, are represented as nodes and patch connectivity or species interactions are depicted as edges that connect the nodes [1, 2]. The way such connections are organized affects system dynamics [3], and thus how the system will respond to changes, such as species loss [4] or changes in the ecological connectivity among patches [5]. Several descriptors have been used to characterize the organization of networks, i.e., the network topology [1, 6]. A recurrent pattern in ecological networks is modularity, also termed compartmentalization, clustering, or community structure [7]. Modules are cohesive groups of highly connected nodes that are loosely connected to other nodes in the network [8, 9].

In ecological networks, modularity will emerge if there are certain groups of individuals, species, or habitat patches that show more interactions among each other than with other groups within the network [10, 11, 12, 6]. This type of modular organization has been found in networks describing different ecological systems, such as resource use by animal populations [13], food webs [14, 15], mutualistic interactions between plant species and their pollinators [9], antagonistic interactions between parasites and their hosts and between plants and their herbivores [16, 17], and the spatial connectivity of metapopulations [18, 12]. Because the degree of modularity determines how dense the connection between different groups of elements in an ecological system is, systems that largely differ in the degree of modularity often differ in their ecological and evolutionary dynamics [10, 9, 19, 5, 20].

Propelled by the relevance of modularity in the dynamics of ecological systems, the detection of modularity became an important aspect of studies that analyze the organization and conservation of ecological networks [5]. Modularity can be measured by different metrics, and one of the most popular modularity metrics in the study of ecological networks is the Q metric (also referred to as M) [8, 21, 9, 16]. For a given partition of a network into modules, Q measures the proportion of edges that connect the nodes within the same module. Because finding the maximal modularity (maximal Q) in a network is an NP-hard problem [22], there is no known algorithm to find the maximum Q in polynomial time. Thus, it is necessary to use sub-optimal optimization approaches to perform this calculation. Heuristic

algorithms can thus be used to approximate a network partitioning with the largest number of edges within the modules and the lowest number of edges between modules to maximize the modularity index (see details of the metrics in Supplementary material Appendix 1). Although module identification is scale-dependent, optimization algorithms can be used to test module consistency across multiple scales, testing the effects of resolution on module detection [23].

Finding the partition with the highest modularity in a large network is often time consuming. Moreover, the analysis of large sets of data and the subsequent testing of the results against theoretical benchmarks that are represented by ensembles formed by thousands of replicates is a common procedure in biology [24, 25]. In this sense, a major constraint in the analysis of the modularity of ecological networks is the lack of programs that allow fast and autonomous computation of modularity for researchers who are not familiar with programming. Here, we introduce the software MODULAR for the computation of the modularity and the identification of modules in multiple complex networks. Many algorithms have been proposed for finding the partition that maximizes the value of Q , and some of these are publicly available. However, to the best of our knowledge, there is no software available that allows multiple uses of different metrics and optimization algorithms in a user-friendly way that accelerates the workflow of the ecologist. MODULAR was developed to allow one to automatically compute the modularity of several input files and to allow one to choose the optimization algorithm that best matches the user’s needs.

MODULAR features

MODULAR was developed in *C* language and uses features from the igraph-0.6 library [26] and the GNU Scientific Library (GSL) [27] (see details in Supplementary material Appendix 2). MODULAR was designed to facilitate and speed up the detection of modules in multiple networks through modularity maximization. To achieve this task, the maximization of modularity is automatically performed for a set of input files containing representations of bipartite networks, such as those depicting species occurrence across islands [6], or unipartite networks, such as spatial networks describing habitat connectivity for a given species [12].

When running MODULAR with a representation of unipartite networks, only the Q metric is available [8]. If the input data files represent bipartite networks, the user can choose between two different modularity metrics: Newman and Girvan’s - Q [8] or Barber’s modularities - Q_B [28], which is a modification of the Q metric for bipartite networks (see details of metrics in Supplementary material Appendix 1). The two options are available for bipartite networks because unipartite indexes have also been used in the ecological literature for bipartite networks (e.g., Olesen *et al.*, 2007; Fortuna *et al.*, 2010 Carstensen *et al.*, 2012).

If the user chooses the traditional Q metric, there are five optimization algorithms that can be used to perform the search for the partition of the network into modules that maximizes the modularity index: (i) fast greedy (FG) [29, 30], (ii) simulated annealing (SA) [21], (iii) spectral partitioning (SP) [31], (iv) a hybrid of simulated annealing and spectral partitioning (Hyb-SP), and (v) a hybrid of simulated annealing and fast greedy (Hyb-FG). The optimization algorithms differ in the method used to search the network partition that maximizes the modularity measurement (see details of MODULAR functioning and optimization algorithms in Supplementary material Appendix 3). Since the running time can vary considerably, the choice of the optimization algorithm becomes particularly important.

To verify whether the modularity found by the maximization of each metric significantly differs from a theoretical benchmark, the user has the option of running two different null models and to specify how many replicates each null model will generate. We included unipartite and bipartite versions of two of the most common null models that are used in studies of ecological networks: (i) the Erdős-Rényi model [32] and (ii) the “null model 2” [33] (see details of the null models in Supplementary material Appendix 4). Because MODULAR can utilize large sets of input data, the user can also test the modularity of the

networks generated by other null models by using those as input data.

MODULAR is an open source software program licensed under the GNU General Public License version 3. It can be downloaded from <http://sourceforge.net/projects/programmodular>. As future directions, we are planning to add new algorithms and optimization methods for the calculation of modularity. In addition, new metrics that analyze modularity at the node level can be added to MODULAR.

Acknowledgments

We thank the igraph developers for promptly solving all of the reported issues in the external interactions of the library. We would also like to thank all of our colleagues who encouraged us to persist in developing this software and who tested it in the earlier stages. We are very thankful to Pedro Jordano for his comments and suggestions. We thank FAPESP for financial support.

References

- [1] M. Pascual and J. A. Dunne, *Ecological networks: linking structure to dynamics in food webs*. Oxford University Press, USA, 2006.
- [2] D. L. Urban, E. S. Minor, E. A. Trembl, and R. S. Schick, “Graph models of habitat mosaics,” *Ecology Letters*, vol. 12, no. 3, pp. 260–273, 2009.
- [3] D. B. Stouffer and J. Bascompte, “Compartmentalization increases food-web persistence,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 9, pp. 3648–3652, 2011.
- [4] J. A. Dunne, R. J. Williams, and N. D. Martinez, “Network structure and biodiversity loss in food webs: robustness increases with connectance,” *Ecology Letters*, vol. 5, no. 4, pp. 558–567, 2002.
- [5] G. S. Cumming, Ö. Bodin, H. Ernstson, and T. Elmqvist, “Network analysis in conservation biogeography: challenges and opportunities,” *Diversity and Distributions*, vol. 16, no. 3, pp. 414–425, 2010.
- [6] D. W. Carstensen, B. Dalsgaard, J.-C. Svenning, C. Rahbek, J. Fjeldsa, W. J. Sutherland, and J. M. Olesen, “Biogeographical modules and island roles: a comparison of wallacea and the west indies,” *Journal of biogeography*, vol. 39, no. 4, pp. 739–749, 2012.
- [7] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006.
- [8] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.
- [9] J. M. Olesen, J. Bascompte, Y. L. Dupont, and P. Jordano, “The modularity of pollination networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 50, pp. 19891–19896, 2007.
- [10] P. R. Guimarães, V. Rico-Gray, P. S. Oliveira, T. J. Izzo, S. F. dos Reis, and J. N. Thompson, “Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks,” *Current Biology*, vol. 17, pp. 1797–803, Oct. 2007.
- [11] M. T. Tinker, P. R. Guimarães Jr, M. Novak, F. M. D. Marquitti, J. L. Bodkin, M. Staedler, G. Benthall, and J. A. Estes, “Structure and mechanism of diet specialisation: testing models of individual variation in resource use with sea otters,” *Ecology Letters*, vol. 15, pp. 475–483, 2012.
- [12] M. Dale and M.-J. Fortin, “From graphs to spatial graphs,” *Annual Review of Ecology, Evolution, and Systematics*, vol. 41, pp. 21–38, 2010.

- [13] M. S. Araújo, P. R. Guimarães Jr, R. Svanbäck, A. Pinheiro, P. Guimarães, S. F. Reis, and D. I. Bolnick, "Network analysis reveals contrasting effects of intraspecific competition on individual vs. population diets," *Ecology*, vol. 89, no. 7, pp. 1981–1993, 2008.
- [14] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, "Compartments revealed in food-web structure," *Nature*, vol. 426, pp. 282–285, 2003.
- [15] S. Allesina and M. Pascual, "Food web models: a plea for groups," *Ecology letters*, vol. 12, no. 7, pp. 652–662, 2009.
- [16] M. A. Fortuna, D. B. Stouffer, J. M. Olesen, P. Jordano, D. Mouillot, B. R. Krasnov, R. Poulin, and J. Bascompte, "Nestedness versus modularity in ecological networks: two sides of the same coin?," *Journal of Animal Ecology*, vol. 79, no. 4, pp. 811–817, 2010.
- [17] M. M. Pires and P. R. Guimarães, "Interaction intimacy organizes networks of antagonistic interactions in different ways," *Journal of The Royal Society Interface*, 2012.
- [18] Ö. Bodin and J. Norberg, "A network approach for analyzing spatially structured populations in fragmented landscape," *Landscape Ecology*, vol. 22, no. 1, pp. 31–44, 2007.
- [19] E. Thébault and C. Fontaine, "Stability of ecological communities and the architecture of mutualistic and trophic networks," *Science*, vol. 329, no. 5993, pp. 853–856, 2010.
- [20] M. Hagen, W. D. Kissling, C. Rasmussen, M. A. De Aguiar, L. L. Brown, D. W. Carstensen, A.-D. Santos, Y. L. Dupont, F. K. Edwards, P. R. Guimarães, *et al.*, "Biodiversity, species interactions and ecological networks in a fragmented world," *Advances in Ecological Research*, vol. 46, pp. 89–210, 2012.
- [21] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, 2005.
- [22] J. Ruan and W. Zhang, "An efficient spectral algorithm for network community discovery and its applications to biological and social networks," in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pp. 643–648, Ieee, 2007.
- [23] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, 2007.
- [24] N. J. Gotelli and G. R. Graves, *Null models in ecology*. Smithsonian Institution Press, 1996.
- [25] N. J. Gotelli, "Research frontiers in null model analysis," *Global Ecology and Biogeography*, vol. 10, no. 4, pp. 337–343, 2001.
- [26] G. Csárdi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006.
- [27] B. Gough, *GNU Scientific Library Reference Manual*. Network Theory Ltd., 3rd ed., 2009.
- [28] M. J. Barber, "Modularity and community detection in bipartite networks," *Physical Review E*, vol. 76, no. 6, p. 066102, 2007.
- [29] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, p. 066111, 2004.
- [30] K. Wakita and T. Tsurumi, "Finding community structure in mega-scale social networks," in *Proceedings of the 16th international conference on World Wide Web*, pp. 1275–1276, ACM, 2007.
- [31] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, no. 3, p. 036104, 2006.
- [32] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.

- [33] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen, “The nested assembly of plant-animal mutualistic networks.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 9383–9387, Aug. 2003.
- [34] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, pp. 671–680, 1983.
- [35] S. P. Borgatti, M. G. Everett, and L. C. Freeman, “Ucinet for windows: Software for social network analysis,” tech. rep., Analytic Technologies, Harvard, MA, 2002.
- [36] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *Journal of Chemical Physics*, vol. 21, pp. 1087–1093, 1953.
- [37] W. Ulrich and N. J. Gotelli, “Null model analysis of species nestedness patterns,” *Ecology*, vol. 88, no. 7, pp. 1824–1831, 2007.

Supplementary material

Appendix 1: Metrics

The metrics used in MODULAR are the Newman and Girvan (2004) [8] metric and its modified version for bipartite networks, which was proposed by Barber (2007) [28]. The first metric is calculated as follows:

$$Q = \sum_{i=1}^{N_M} \left[\frac{E_i}{E} - \left(\frac{k_i}{2E} \right)^2 \right], \quad (1)$$

where N_M is the number of modules, E_i is the number of links in module i , E is the number of links in the complete network, and k_i is the sum of the degrees of the nodes within module i . Thus, a good partition has many links inside the modules and as few links as possible between the modules.

The bipartite version of the metric is calculated as follows:

$$Q_B = \sum_{i=1}^{N_M} \left[\frac{E_i}{E} - \left(\frac{k_i^C \cdot k_i^R}{E^2} \right) \right], \quad (2)$$

where k_i^C is the sum of the degrees of the nodes within module i that belong to set C and k_i^R is the sum of the degrees of the nodes within module i that belong to set R .

Appendix 2: Language, Libraries, and Algorithms

We implemented the MODULAR software in *C*. Our code depends on two libraries: igraph and the GNU Scientific Library (GSL). The igraph library provides data structures for graph representation, manipulation functions and functions for community structure detection. The GSL library is mainly used to generate the random numbers that are needed by the optimization methods. By implementing a set of functions utilizing those two libraries, we were able to develop an easy-to-use software program for the detection of modules with maximum modularity. The igraph library utilized is the 0.6 version, which supports UCINET's DL file format. We used the *igraph_read_graph_dl* function to read the network into a graph data structure. In addition, we used igraph data structures called *membership vectors* to represent the modules of the network. The SA [34] method was implemented using a data structure from GSL that contains a number of running parameters, such as the initial temperature and its damping factor. The *fast greedy* [29] and the *spectral partitioning* [31] methods are functions called from the igraph library. These two methods are faster than simulated annealing, but their searches are based on local decisions and thus eventually lead to lower modularities than those found with SA. In MODULAR, we combined the speed of these two methods to find a reasonable solution with the broader search that is provided by the simulated annealing method by creating two hybrid methods. The first hybrid approach (*Hyb-SP*) first runs the SP method, and its module configuration output is utilized as the input for the SA method using a reduced initial temperature. The same technique is applied in the second hybrid method, except that the FG method is used instead of the SP method (*Hyb-FG*).

Appendix 3: MODULAR Functioning

The basic MODULAR functioning is represented in Figure A11. The set of input files can include two kinds of files: networks represented as matrices in text files (extension .txt) or list of interactions, such as those used by the UCINET program (extension .dl) [35]. Both kinds of input files can contain representations of *bipartite networks* or *unipartite networks*. If the input data are text files containing

matrices, each *.txt* file is a binary matrix in which the columns are separated by tabulation or a space. The user must not include any row or column labels in the input file. The representation of unipartite networks should contain square matrices in which the rows i and columns i depict the same element in the $N \times N$ matrix. These matrices are symmetrical, and the program will read only the input data from the upper triangular part.

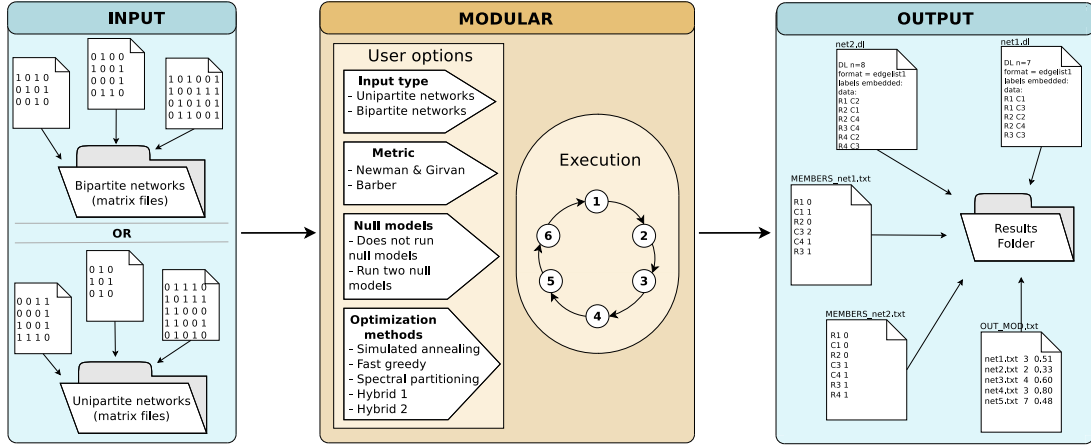


Figure 1. Design of the MODULAR software. The program design can be divided into three steps. The first step is the *input* file preparation. The user may choose between representations of bipartite or unipartite networks. The second step includes the running of MODULAR. After the user answers some questions, MODULAR will execute six steps (see the execution explanation in the text). Finally, MODULAR will produce *output* files with the results.

The current version of the program is not intended to be used with directed networks. If the matrices in the text files represent bipartite $N \times M$ networks, the rows and columns represent different sets of elements, and edges exist only between elements of the two different sets. For example, the rows can represent islands and the columns can represent species in a network describing the species co-occurrence across islands [6]. In contrast, if the input data are UCINET files, the representation of unipartite networks should follow the UCINET file formats with the *format* field set to contain the format type *edgelist1* followed by an edge list of the node pairs. The representation of bipartite networks should set the *format* field to *edgelist2*, followed by an edge list of the node pairs. The first node of the pair is an element of one set, and the second node is an element of the other set. Therefore, the first option that MODULAR gives the user is the selection of different types of input files that represent bipartite and unipartite networks.

The second option given to the user regards the modularity metric: Newman and Girvan's - Q [8] or Barber's modularities - Q_B [28] (see the Appendix 1 for details). If the user chooses the traditional Q metric, there are five optimization algorithms that can be used: (i) fast greedy (FG) [29, 30], (ii) simulated annealing (SA) [21], (iii) spectral partitioning (SP) [31], (iv) a hybrid of simulated annealing and spectral partitioning (Hyb-SP), and (v) a hybrid of simulated annealing and fast greedy (Hyb-FG). The optimization algorithms differ in the method used to search the network partition that maximizes the modularity measurement (objective function).

The FG method searches the maximum modularity by only accepting partitions that exhibit a higher modularity than the previous partition. It is a greedy optimization method. It was proposed by Clauset *et. al* (2004) for the analysis of modularity and improved by Wakita & Tsurumi (2007). The method is summarized below:

FG optimization:

- 1) Compute the modularity of the current partition of the network.
- 2) Join the communities and accept the new partition that provides the highest modularity.
- 3) Stop when no partition joining can increment the modularity.

The SA method searches the maximum modularity while trying to avoid getting stuck at local maxima [34]. This time-consuming method first partitions the network into modules with one node each, thus generating an initial configuration of N modules of size 1. The method then randomly moves nodes between modules and computes the resulting modularity of the new module configuration. If the new modularity is higher than that obtained with the previous module configuration, the algorithm takes the new module configuration as the current solution. However, if the new modularity is lower than that obtained with the previous module configuration, the algorithm accepts it as the new solution with a given probability [36], thus simulating a system of particles in which the energy levels follow a Boltzmann distribution. Using this technique, the algorithm avoids getting trapped at local maxima by moving through sub-optimal solutions and potentially converging to a maximal point [34]. This avoidance is possible because of the *temperature* parameter, which is directly linked to the probability of accepting new solutions. The temperature decreases gradually during the process, thereby decreasing the probability to accept worse solutions (lower modularity values). The steps are repeated until the SA temperature reaches a threshold or the algorithm undergoes a predetermined number of iterations without any changes in the solution. If the user chooses to maximize Barber’s metric, the simulated annealing method is the only search method available in the current version of the program. The method is summarized below:

SA optimization:

- 1) Compute the modularity in the current partition of the network.
- 2) Generate a new module partitioning by randomly exchanging nodes and merging/splitting modules, compute the modules, and determine the modularity.
- 3) Accept the new partition if the modularity increases. If the modularity decreases, accept the new partition with a given probability based on the difference between the new and the previous modularity values and according to the chosen temperature parameter used to run the SA.
- 4) Stop if the solution does not change after a predetermined number of iterations or when the minimum temperature chosen to run the SA is reached.

The SP method is based on matrix spectra. This fast method divides the network into sequential groups according to the entries in the leading eigenvector [29].

SP optimization:

- 1) Establish the modularity matrix (matrix B, see Newman 2006).
- 2) Divide the network into two groups according to the entries in the leading eigenvector.
- 3) Repeat the process until none of the sub-matrices have a positive eigenvalue.

In MODULAR, we introduced two additional hybrid methods that combine the speed of the SP and FG approaches with the larger search space of simulated annealing. In the hybrid approaches, the SA input is the output of the faster method (SP or FG). Consequently, the optimization can achieve better

results. After these steps, the user can choose to verify the modularity found by the maximization against two different null models.

Based on the options above, MODULAR iteratively executes up to six steps (“Execution”, in Figure A1) for each *.txt* or *.dl* file present in the matrices folder specified by the user:

For each *.txt* or *.dl* file in the specified folder, do:

- 1) Read the matrix A or the edge list from the input file according to the specified type (representation of a bipartite or unipartite network);
- 2) If the input file is a matrix, remove all null rows and null columns present in the matrix to generate a new matrix A' ;
- 3) If the input file is a matrix, generate a *.dl* UCINET [35] file with the network represented by matrix A' ;
- 4) Read the *.dl* file and load the network into the computer memory using data structures that can be manipulated by the *igraph* library functions;
- 5) Maximize the modularity of the network by running the method selected by the user (SA, SP, FG, Hyb-SP, or Hyb-FG);
- 6) Generate the theoretical networks according to the null model (if requested by the user) and run step 5 for each generated network.

A set of output files is generated after these steps. There are three main output files. If the input files are named *MatrixName.txt*, MODULAR generates a file with the list of interactions named *MatrixName.dl* in the UCINET edgelist1 format. The first line indicates the number of nodes in the network. If MODULAR was used to analyze bipartite network representations, the rows will be named R followed by a number (for example, $R2$ for the second row of the input file), and the columns will be named C followed by a number (for example, $C3$ for the third column of the input file). If the data set is composed of unipartite network representations, the lines will be named L followed by a number (for example, $L2$ for the second line of the input file). In both cases, the number refers to the order found in the input file.

A set of output files is generated after these steps. The main output is the *OUT_MOD.txt* file. This file contains, for each input file, the value of the modularity metric and the number of modules found by the metric and algorithm chosen. If the user chooses to run the null models, the *OUT_MOD.txt* file will also contain the proportion of null matrices with modularity higher than that in the input file. An output file named *MEMBERS_MatrixName.txt* is generated for each input file named *MatrixName.txt*. This file has two columns: the first column has the labels of the matrix lines from the *.dl* file and the second column indicates the module to which each line belongs. If null model analysis is performed, there will be two additional outputs: *OUT_1_MatrixName.txt*, which includes the modularities of null model 1, and *OUT_2_MatrixName.txt*, which includes the modularities of null model 2. Both files have as many lines as the number of null model runs chosen by the user.

Appendix 4: Null models

The null models implemented in MODULAR are the Erdős-Rényi model [32] and “null model 2” [33]. In the Erdős-Rényi model, each pair of nodes has the same probability of being connected by an edge. Thus, the probability of a pair (i, j) to be linked is given by

$$P(i, j) = \frac{E}{R \cdot C}, \quad (3)$$

where E is the number of edges in the network. If the network is bipartite, R and C are the number of nodes in each set of the network. For unipartite networks, $R = C$ and the probability of a pair (i, j) to be linked is given by

$$P(i, j) = \frac{2E}{R \cdot (R - 1)}. \quad (4)$$

In “null model 2” [33], the probability of a pair being connected by an edge is proportional to the number of edges that the nodes have. Thus, the probability of a pair (i, j) to be linked is given by

$$P(i, j) = \frac{1}{2} \left(\frac{k_{i \in R}}{C} + \frac{k_{j \in C}}{R} \right), \quad (5)$$

where k_i is the number of edges of node i in the R partition and k_j is the number of edges of node j in the C partition. For adjacency matrices, $R = C$.

For additional information on the potential uses of null models, please refer to the study published by Ulrich and Gotelli [37].